Available online at http://scik.org
J. Math. Comput. Sci. 2 (2012), No. 1, 82-90
ISSN: 1927-5307

A MODIFIED PRP CONJUGATE GRADIENT METHOD

JINKUI LIU*, BENXIU LI

School of Mathematics and Statistics, Chongqing Three Gorges University, Wanzhou, P.R.China

Abstract. In this paper, a class of new conjugate gradient method with variable parameters is proposed to solve unconstrained optimization problems on the base of PRP method. Under the strong Wolfe line searches, we proved the global convergence of the new method without the given sufficient descent condition. Many numerical experiments show that the new method is very efficient.

Keywords: Unconstrained optimization; Conjugate gradient method; Strong Wolfe line searches; Sufficient descent property; Global convergence.

2000 AMS Subject Classification: 90C26; 65H10

1. Introduction

Consider the following unconstrained optimization problem

$$\min_{x \in R^n} f(x), \tag{1}$$

*Corresponding author

E-mail addresses::liujinkui2006@126.com(J.Liu), benxiuli@126.com(B. Li)

This work was supported by The Nature Science Foundation of Chongqing Education Committee (KJ091104) and The Youth Project Foundation of Chongqing Three Gorges University (11QN-26).

Received December 2, 2011

where$f: R^n \to R$ is smooth and its gradient $g(x) = \bigtriangledown f(x)$ is available.

Conjugate gradient methods for solving (1) are iterative formulas of the form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

$$d_k = \begin{cases} -g_k, & for\ k = 1, \\ -g_k + \beta_k d_{k-1}, & for\ k \geq 2. \end{cases} \tag{3}$$

where $g_k = \bigtriangledown f(x_k)$, x_k is the current iterate; α_k is a positive scalar and called the steplength which is determined by some line searches;d_k is the search direction , and β_k is a scalar. There are many ways to select β_k, and some well-known formulas are given by

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}(Fletcher and Reeves (FR)[1]),$$

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2}(Polak - Ribiére - Polyak[2],[3]),$$

where $||.||$ is the Euclidean norm. In the convergence analysis and implementations of conjugate gradient methods, one often requires the inexact line search such as the Wolfe line search or the strong Wolfe line search.

The Wolfe line searches require α_k satisfying:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \tag{4}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \tag{5}$$

where $0 < \delta < \sigma < 1$. The strong Wolfe line searches require satisfying (4) and

$$|g(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k. \tag{6}$$

where $0 < \delta < \sigma < 1$.

In the exact line search, [4] proved that the FR method was global convergence for general non-convex function, [5] generalized the results to the case of inexact line search; but even in the exact line search, the PRP algorithm also has no overall convergence. In the Wolfe line search, [6] proved the PRP algorithm is global convergence under assuming the sufficient descent condition. Gilbert and Nocedal [7] proved that the conjugate gradient method with $\beta_k = max\{0, \beta_k^{PRP}\}$ converged globally, where the Wolfe line searches and sufficient descent condition were satisfied. Thus, the selection of the parameters β_k for research on the conjugate gradient algorithm is important.

In this paper, we will propose a new parameter β_k, and through a simple method to prove the global convergence of the new method with the strong Wolfe line searches.

2. The sufficient descent property

Throughout this paper we make the following assumptions on the objective function.

Assumption (H):(i) The level set $N = \{x \in R^n | f(x) \leq f(x_1)\}$ is bounded, where x_1 is the starting point.

(ii) In some neighborhood W of N, the objective function is continuously differentiable, and its gradient is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \ \forall \ x, y \in W. \tag{7}$$

Algorithom 2.1:

Step 1: Data: $x_1 \in R^n$, $\varepsilon \geq 0$. Set $d_1 = -g_1$, if $||g_1|| \leq \varepsilon$, then stop.

Step 2: Compute α_k by the strong Wolfe line searches.

Step 3: Let $x_{k+1} = x_k + \alpha_k d_k$, $g_{k+1} = g(x_{k+1})$, if $||g_{k+1}|| \leq \varepsilon$, then stop.

Step 4: Compute

$$\beta_{k+1} = \begin{cases} \frac{\|g_{k+1}\|^2 - \rho|g_{k+1}^T g_k|}{u(g_{k+1}^T d_k)^2 + \|g_k\|^2}, if \|g_{k+1}\|^2 \geq |g_{k+1}^T g_k| \\ 0, \ else. \end{cases} \tag{8}$$

where $\rho \in [0, 1]$,$u \geq 0$; and d_{k+1} is computed by (3).

Step 5: Set $k = k + 1$, go to step 2.

Theorem 2.1 Consider any method (2) and (3), where α_k satisfies the strong Wolfe line search and β_k is computed by (8), then for all $k \geq 1$, we have

$$\frac{1 - 2\sigma + \sigma^k}{1 - \sigma} \leq -\frac{g_k^T d_k}{\|g_k\|^2} \leq \frac{1 - \sigma^k}{1 - \sigma}. \tag{9}$$

Proof. The conclusion can be proved by induction. Since $-\frac{g_1^T d_1}{\|g_1\|^2} = 1$, (9) holds for $k = 1$. Now we assume that (9) is true for $k - 1$ and $g_k \neq 0$. We need to prove that (9) holds for

k. From (3), we have

$$-\frac{g_k^T d_k}{\|g_k\|^2} = 1 - \beta_k \frac{g_k^T d_{k-1}}{\|g_k\|^2}. \tag{10}$$

From (8), we have

$$0 \le \beta_k \le \frac{\|g^T\|^2}{\|g_{k-1}\|^2}. \tag{11}$$

Then from (10) and (11), we have

$$1 - \frac{|g_k^T d_{k-1}|}{\|g_{k-1}\|^2} \le -\frac{g_k^T d_k}{\|g_k\|^2} \le 1 + \frac{|g_k^T d_{k-1}|}{\|g_{k-1}\|^2}. \tag{12}$$

From (6) and (12), we have

$$1 + \sigma \frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \le -\frac{g_k^T d_k}{\|g_k\|^2} \le 1 - \sigma \frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}. \tag{13}$$

Using the induction hypothesis and the first inequality of (13), we have

$$-\frac{g_k^T d_k}{\|g_k\|^2} \ge 1 - \sigma \frac{1-\sigma^{k-1}}{1-\sigma} = \frac{1-2\sigma+\sigma^k}{1-\sigma}.$$

Using the induction hypothesis and the first inequality of (13), we also have

$$-\frac{g_k^T d_k}{\|g_k\|^2} \le 1 + \sigma \frac{1-\sigma^{k-1}}{1-\sigma} = \frac{1-\sigma^k}{1-\sigma}.$$

This shows that (8) holds for k as well. Therefore the theorem is completed.

3. Global convergence of the new method

In the following we state a lemma which was shown by Z.F.Li, J.Chen and N.Y.Deng (see [8]).

Lemma 3.1[8] Suppose Assumption (H) holds. Consider any iteration in the form (2)-(3), where d_k satisfies $g_k^T d_k < 0$ for $k \in N^+$ and α_k satisfies the strong Wolfe line searches. Then

$$\sum_{k \ge 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{14}$$

Theorem 3.1 Suppose that Assumption (H) holds. Consider any method of the form (2)-(3), where β_k is computed by (8), and where α_k satisfies the strong Wolfe line searches. Then,

$$\liminf_{k\to+\infty} \|g_k\| = 0.$$

Proof. We assume the conclusion is not true, then there exists a constant $\gamma > 0$ such that for all

$$\|g_k\|^2 \geq \gamma. \tag{15}$$

From (3), we have

$$\|d_k\|^2 = \|g_k\|^2 + \beta_k^2\|d_{k-1}\|^2 - 2\beta_k g_k^T d_{k-1}.$$

Then from (11), we have

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leq \frac{1}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + 2\frac{|g_k^T d_{k-1}|}{\|g_k\|^2\|g_{k-1}\|^2}. \tag{16}$$

Set $t_k = \frac{\|d_k\|^2}{\|g_k\|^4}$. From (16), we have

$$t_k \leq t_{k-1} + \frac{1}{\|g_k\|^2}(1 + 2\frac{|g_k^T d_{k-1}|}{\|g_{k-1}\|^2}). \tag{17}$$

From (6) and (9), we have

$$\frac{|g_k^T d_{k-1}|}{\|g_{k-1}\|^2} \leq -\sigma\frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \leq \sigma\frac{1-\sigma^{k-1}}{1-\sigma} \leq \frac{\sigma}{1-\sigma}. \tag{18}$$

Then from (17) and (18), we have

$$t_k \leq t_{k-1} + \frac{1+\sigma}{1-\sigma}\cdot\frac{1}{\|g_k\|^2}. \tag{19}$$

As $t_1 = \frac{1}{\|g_1\|^2}$, from (19), we have

$$t_k \leq \frac{1+\sigma}{1-\sigma}\sum_{i=1}^{k}\frac{1}{\|g_i\|^2}. \tag{20}$$

From (15) and (20), we have

$$t_k^{-1} \geq \frac{(1-\sigma)\gamma}{(1+\sigma)k}. \tag{21}$$

Then

$$\sum_{k\geq 1} t_k^{-1} = +\infty. \tag{22}$$

From (14), (9) and (20), we have

$$+\infty > \sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \sum_{k\geq 1} [(\frac{(1-2\sigma+\sigma^k}{1-\sigma})^2 t_k^{-1}] \geq (\frac{1-2\sigma}{1-\sigma})^2 \sum_{k\geq 1} t_k^{-1}.$$

Obviously, we have

$$\sum_{k\geq 1} t_k^{-1} < +\infty. \tag{23}$$

From (21) and (22), the assumption does not hold, so the conclusion holds.

4. Numerical results

In order to investigate the numerical of the proposed new conjugate gradient algorithm, we use the MATLAB programming to test the Algorithm 2.1on problems in [9]. The termination condition is $\|g_k\| \leq 10^{-6}$, or It-max >9999. It-max denotes the maximal number of iterations. We test the following four conjugate gradient methods:

New1: β_k is computed by (8), $\rho = 1, u = 0, \delta = 0.01, \sigma = 0.1$.

New2: β_k is computed by (8), $\rho = 0.25, u = 0.2, \delta = 0.01, \sigma = 0.1$.

New3: β_k is computed by (8),$\rho = 0.25, u = 1, \delta = 0.01, \sigma = 0.1$.

New4: β_k is computed by (8), $\rho = 1, u = 1, \delta = 0.01, \sigma = 0.1$.

The numerical results of our tests are reported in Table 1. The column " Problem " represents the problem's name in [9]. The detailed numerical results are listed in the form NI/NF/NG, where NI, NF, NG denote the number of iterations, function evaluations, and gradient evaluations respectively. "Dim" denotes the dimension of the test problems. In order to rank the average performance of all above conjugate methods, one can compute the total number of function and gradient evaluation by the formula

$$N_{total} = NF + l * NG, \tag{24}$$

where l is some integer. According to the results on automatic differentiation [10,11], the value of l can be set to 5, i.e.

$$N_{total} = NF + 5 * NG. \tag{25}$$

That is to say, one gradient evaluation is equivalent to five function evaluations if automatic differentiation is used.

By making used of (24), we compare the Algorithm 2.1 with PRP$^+$ method as follows: for the ith problem , compute the total number of function evaluations and gradient evaluations required by the evaluated method j(denoted shortly by $New(j)$) and PRP$^+$ method by formula (24), and denote them by $N_{total,i}(New(j))$ and $N_{total,i}(PRP^+)$, then calculate the ratio

$$\gamma_i(New(j)) = \frac{N_{total,i}(New(j))}{N_{total,i}(PRP^+)}.$$

able 1: The results for Algorithm 2.1 and PRP$^+$ method

Problem	Dim	PRP$^+$	New1	New2	New3	New4
ROSE	2	24/114/90	29/140/113	33/136/115	29/119/101	23/111/87
FROTH	2	11/72/56	11/69/53	10/71/55	9/61/45	12/87/70
BADSCP	2	45/232/209	26/218/202	22/187/172	26/218/200	22/172/159
BADSCB	2	12/76/69	13/122/110	30/250/225	36/289/266	37/334/310
BEALE	2	13/58/45	17/62/51	17/71/58	16/65/50	14/55/44
JENSAM	2(m=6)	10/37/23	10/39/24	13/48/31	14/49/31	13/49/32
HELJX	3	65/181/156	39/127/106	39/125/103	35/110/93	40/129/114
BARD	3	16/64/51	16/53/42	20/79/65	21/80/67	27/85/69
GAUSS	3	3/7/4	3/8/6	3/8/6	3/8/6	3/8/6
SING	4	113/379/328	58/220/142	70/282/248	98/353/307	59/244/208
WOOD	4	118/357/304	47/208/174	51/187/152	62/225/186	45/177/147
KOWOSB	4	93/269/240	57/217/194	34/130/114	36/140/122	52/193/171
WATSON	3	42/104/87	15/62/48	14/59/44	9/35/22	17/63/47
	5	133/374/330	88/298/254	62/212/181	70/243/210	80/263/223

SINGX	500	105/342/297	61/242/210	60/242/210	67/238/203	66/278/240
	1000	198/693/595	61/243/212	75/313/273	47/191/164	76/292/255
TRIG	100	58/120/113	63/146/133	61/136/130	61/136/130	66/148/135
	200	64/135/128	59/136/127	57/111/108	57/111/108	59/135/125
BV	500	1645/2889/2888	124/309/279	97/184/180	97/184/180	161/402/364
	1000	147/251/250	17/33/29	42/69/68	42/69/68	17/33/29
TRID	500	35/78/74	31/71/57	34/76/71	34/75/70	35/80/61
	1000	34/76/72	35/79/75	34/77/73	34/77/73	33/75/70

The geometric mean of these ratio for method $New(j)$ over all the test problems is defined by

$$\gamma(New(j)) = (\prod_{i\in S} \gamma_i(New(j)))^{\frac{1}{|S|}},$$

where S denotes the set of the test problems, and $|S|$ denotes the number of elements in S. One advantage of the above rule is that, the comparison is relative and hence does not be dominated by a few problems for which the method requires a great deal of function evaluations and gradient functions.

Table2: Relative efficiency of the Algorithm 2.1 and PRP$^+$ method

PRP$^+$	New1	New2	New3	New4
1	0.7132	0.7421	0.6891	0.7994

According to the above rule, it is clear that $\gamma(PRP^+) = 1$. From Table 2, we can see that average performance of Algorithm 2.1 is much better than the PRP^+ method. Through a large number of numerical tests, we found that the numerical results and convergence property of the new algorithm is better when the parameters $\rho \in (0, 1]$ and $u \in (0, 1]$.

References

[1] R. Fletcher, C. Reeves. Function minimization by conjugate gradients, Computer Journal 7(1964): 149 -154.

[2] KE. Polak, and G. Ribire. Note sur la xonvergence de directions conjugees, Rev Francaise informat Recherche Operatinelle 3e Annee 16(1969):35 - 43.

[3] B.T.Polak. The conjugate gradient method in extreme problems, USSR Comput. Math. Math. Phys. 9(1969) 94 -112.

[4] Powell, M, J. D., Nonconvex Minimization Calculation and the Conjugate Gradient Method, In Numerical Analysis, Dundee, 1983, Griffiths, D.E.ed.

[5] Al-Baali M. Descent property and global convergence of the Fletcher-Reeves method with inexact line search[J]. IMA J Numer Anal, 1985,(5): 121 - 124.

[6] Yuan Y. Nonlinear Conjugate Gradient Method (in Chineses). Shanghai:Shanghai Scientific and Technical Publishers,2000,10.

[7] Gilbert, J.C. and Nocedal, J. Global convergence properties of conjugate gradient methods for optimization [J]. SIAM. J. Optimization, 1992, 1:21 - 42.

[8] Z.F.Li, J.Chen, N.Y.Deng, Convergence properties of conjugate gradient methods with Goldstein line searchs, J.China Agric.Univer. I (4) (1996) 15-18.

[9] J.J.More, B.S.Garbow, K.E.Hillstrome, Testing unconstrained optimization software, ACM Trains. Math. Software 7(1981) 17-41.

[10] Y. Dai, Q. Ni, Testing different conjugate gradient methods for large-scale unconstrained optimization, Jornal of Computational Mathematics 213(2003)11-320.

[11] A. Griewank, On automatic differentiation, in: M. Iri, K. Tannabe(Eds.), Mathematical Programming: Recent Developments and Applications, Kluwer Academic dPublishers, 1989, pp. 84-108.